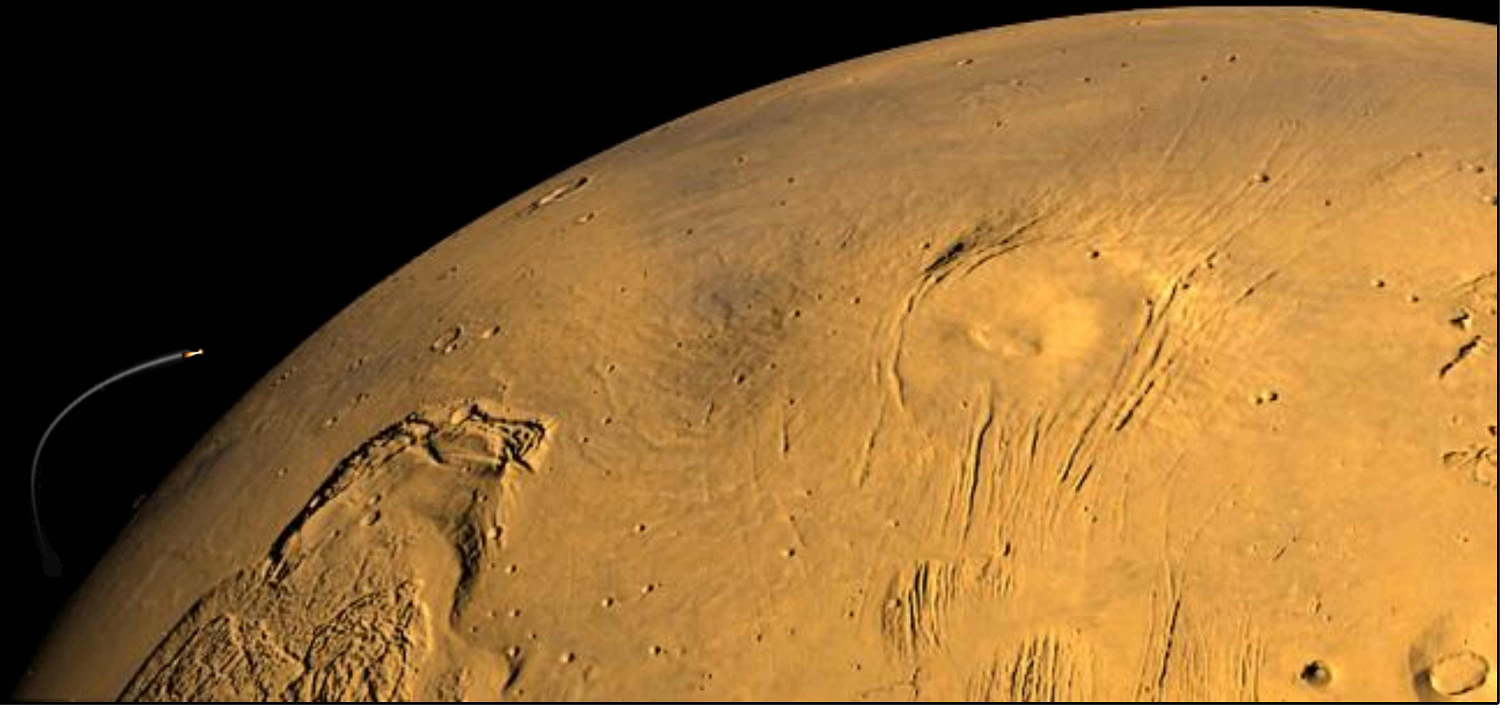
Mars Ascent Vehicle



Jet Propulsion Laboratory
California Institute of Technology

MARSHALL
SPACE FLIGHT CENTER

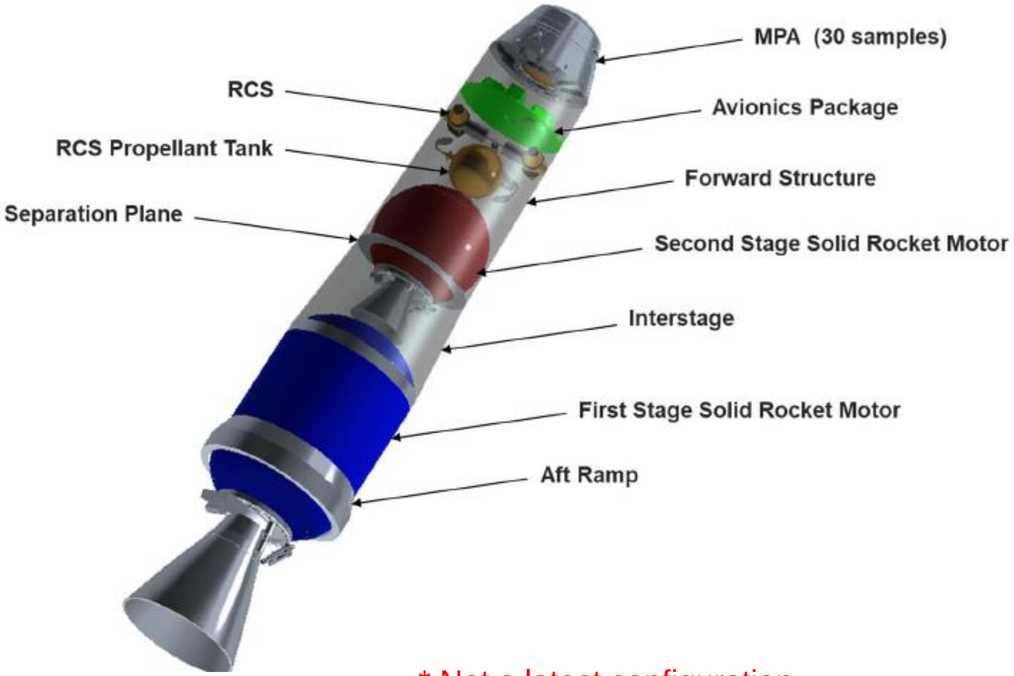
An Online Reinforcement Learning Controller Design for Mars Ascent Vehicle

February 5th, 2022
2021 AAS GN&C Conference
Breckenridge, CO, USA

Han Woong (Brian) Bae, PhD
Control System Design & Analysis

Mars Ascent Vehicle Overview

- Mars Ascent Vehicle (MAV) is a part of Mars Sample Return Campaign.
  – Goal: Return the Martian soil to Earth.
- MAV consists of multiple stages
  – Focus on first stage Solid Rocket Motor pitch and yaw maneuver with TVC
- Two control system design challenges:
  – Abruptly changing vehicle parameters.
  – Disadvantageous Martian setup.



* Not a latest configuration

Source: Yaghoubi, Schnell. "Mars Ascent Vehicle Solid Propulsion Configuration", 2020 IEEE Aerospace Conference, March 2020

Problems

Conditions:

- ✓ Knowledge of atmospheric state
- ✓ Physical access to hardware
- ✓ Instant communication
- ✓ Safe to board human

Conditions:

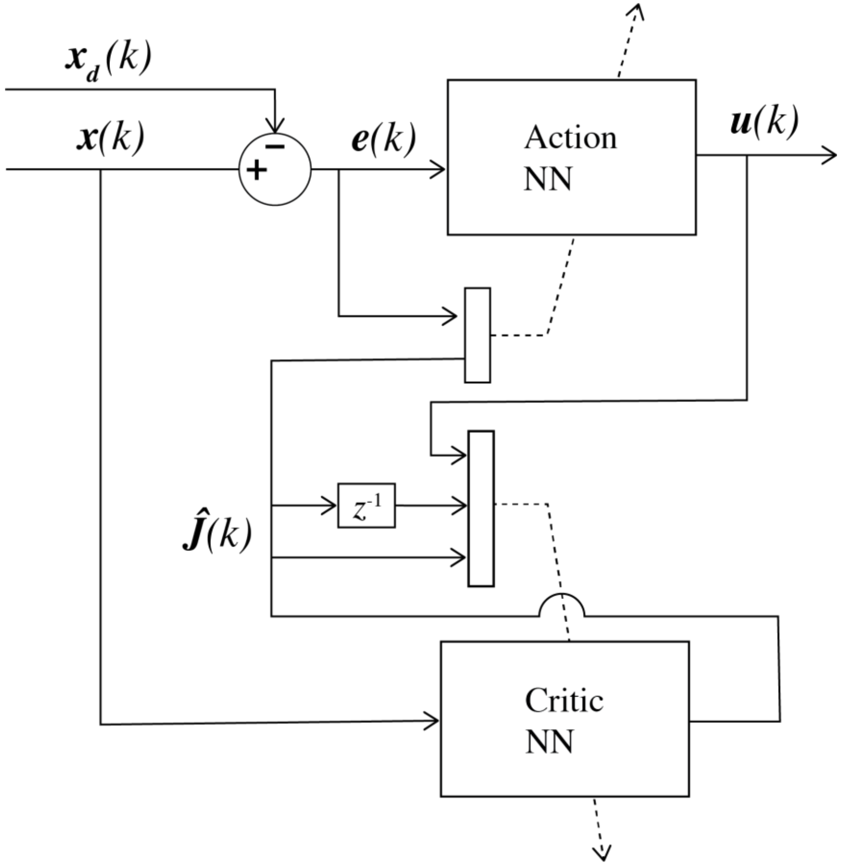
- ✗ Knowledge of atmospheric state
- ✗ Physical access to hardware
- ✗ Instant communication
- ✗ Safe to board human

Solution: Automation

- The long-term solution is to automate the vehicle.
  - Quickly adapt to off-nominal conditions.
  - Optimize performance in-flight.
  - No Deep Space Network (DSN) communication needed.
- AI/ML will lead to increase in automation as well as GN&C performance.[2]
- Online reinforcement learning control by Yang et al is proposed for this study[3]
  - No prior training required.
  - Superior adaptability on-flight.
  - Robust against disturbances.

Overview of ORL controller [1]

- The online reinforcement learning control involves two neural networks (NN):
  - The action NN (control law) applies short term performance measure based on output error terms.
  - The critic NN (performance cost function) evaluates the performance of the action based on the long-term cost function



Overview of ORL controller [2]

- Action NN Law and Weight Update
  - The action NN law is given by:

$$\vec{u}(k) = \hat{w}_a(k)^T \vec{\phi}'_a(v_a^T\, \vec{s}\,(k))$$

  - Action NN weight update is given by:

$$\hat{w}_a(k+1) = \hat{w}_a(k) + \alpha_a \vec{\phi}'_a\left(v_a^T\, \vec{s}\,(k)\right)\left(\vec{e}(k+1) - l_1\vec{e}(k) + \hat{\vec{J}}(k)\right)^T$$

- Critic NN and Weight update
  - Let cost function $J(k)$ be defined by:

$$\vec{J}(k) = \sum_{i=0}^{\infty} \gamma^i \vec{r}(k+i)$$

$$r_i(k) = Q_i|e_i(k)|e_i(k) + R_i|u_i(k)|u_i(k), i = 1, 2, \ldots, 6$$

  - The cost function can be estimated with the critic NN law given by

$$\hat{\vec{J}}(k) = \hat{w}_c^T(k)\vec{\phi}_c(v_c^T\vec{x}(k))$$

  - Where critic NN weight update is given by:

$$\hat{w}_c^T(k+1) = \hat{w}_c^T(k) - \alpha_c\gamma\vec{\phi}'_c(\vec{x}(k))(\gamma\hat{\vec{J}}(k) - \vec{r}(k) - \hat{\vec{J}}(k-1))$$

Simulation Set up

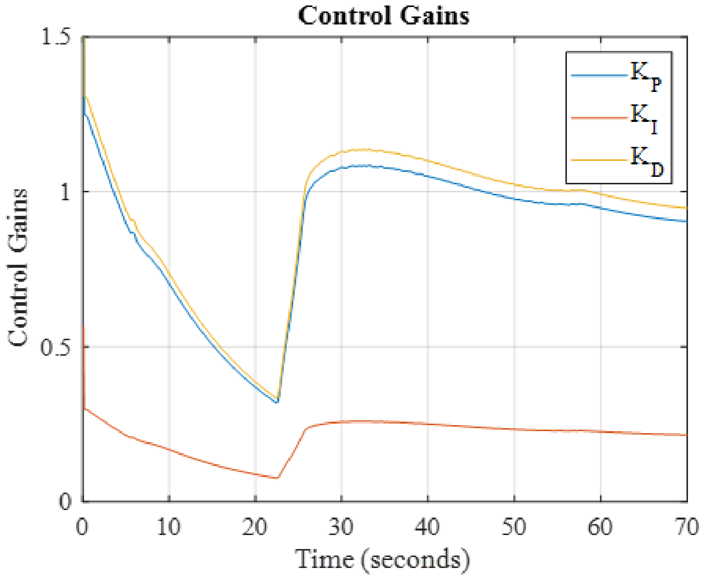
- The ORL controller is set up for the pitch and yaw maneuver on the first stage.
- TVC internal control assumed second-order dynamics with bandwidth of 10 Hz and damping coefficient of 0.707.
- The simulation includes sensor noises, sensor bias, navigation latencies and guidance steering.
- The ORL controller is compared with PID control and gain-scheduled pole-placed PID control.
- Two test cases are considered:

Table 1. Test Case Descriptions

Case #	Description
Case 1	Nominal
Case 2	High external disturbance

Simulation Set up

- The optimized constant gains for PID are 0.5, 0.2, and 0.2 for K_p, K_d, and K_I, respectively.
- GS-PP-PID gains are variable and obtained through pole-placement using the lookup tables of the parameters and accelerometer readings.
  - Desired control bandwidth and damping coefficient are 1 Hz and 1, respectively.
- The tuned gains for ORL is shown in Table 2.

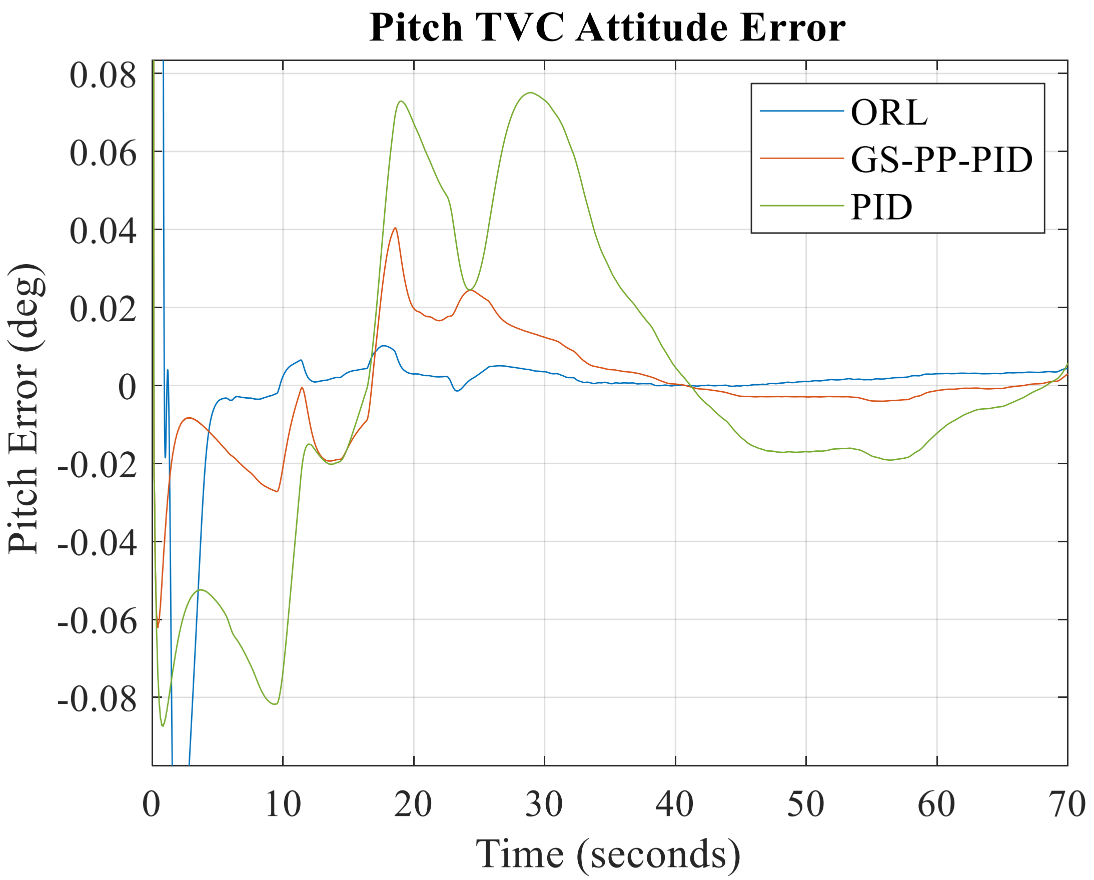


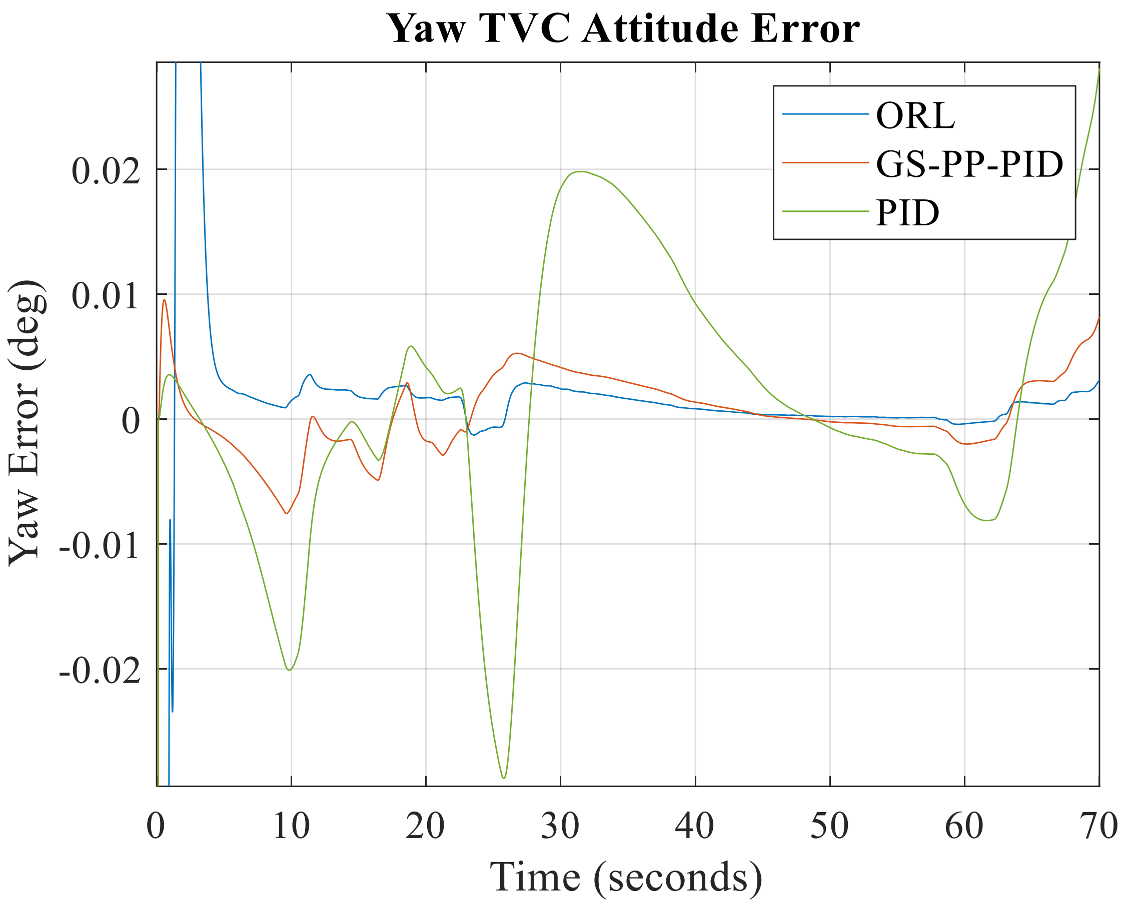
GS-PP-PID gains throughout the first stage TVC powered flight.

Table 2. Control Gain Values

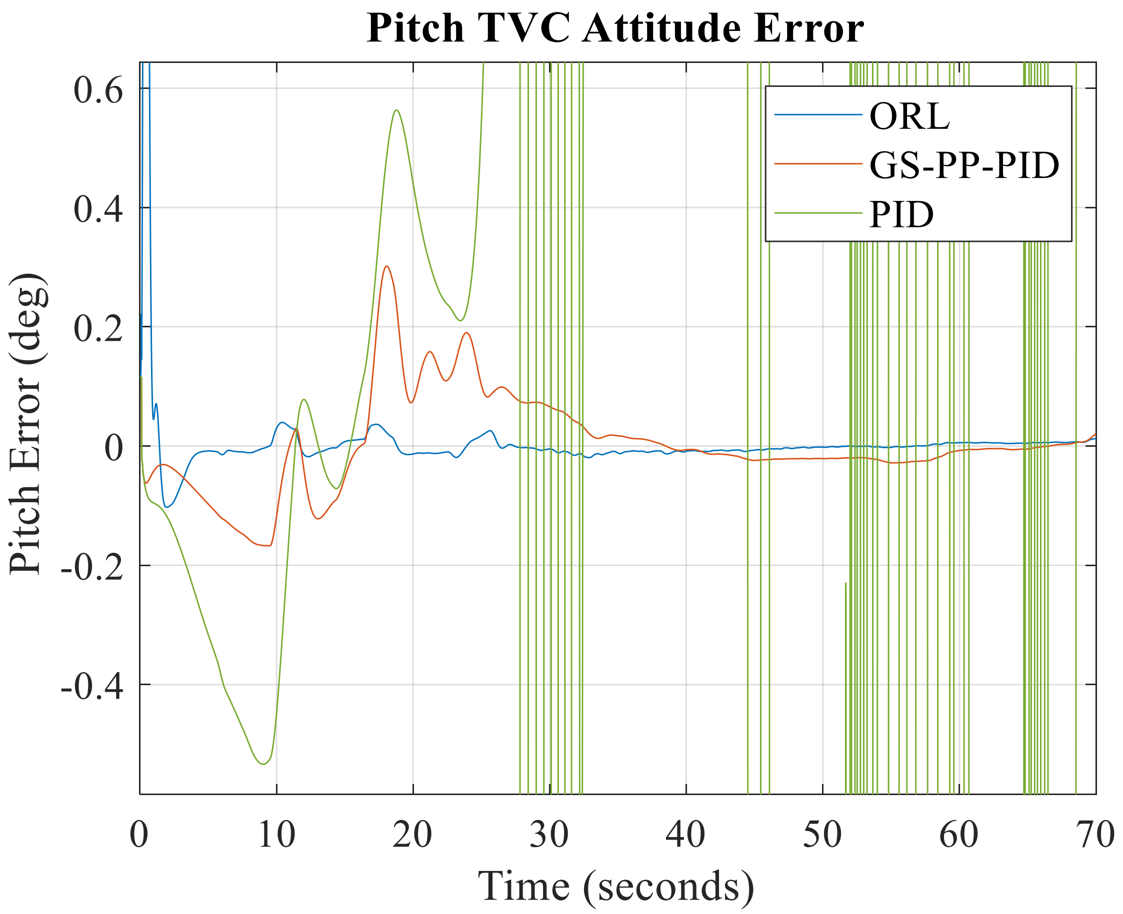
Parameter	Value
$\vec{Q}$	[1 1 5 5 7 7]
$\vec{R}$	$[3\ 3\ 0\ 0\ 0\ 0] \times 10^5$
α_a	$[34\ 34\ 16\ 16\ 2.1\ 2.1] \times 10^{-4}$
α_c	1×10^{-4}
l_1	0.2
γ	0.5
p	1×10^6

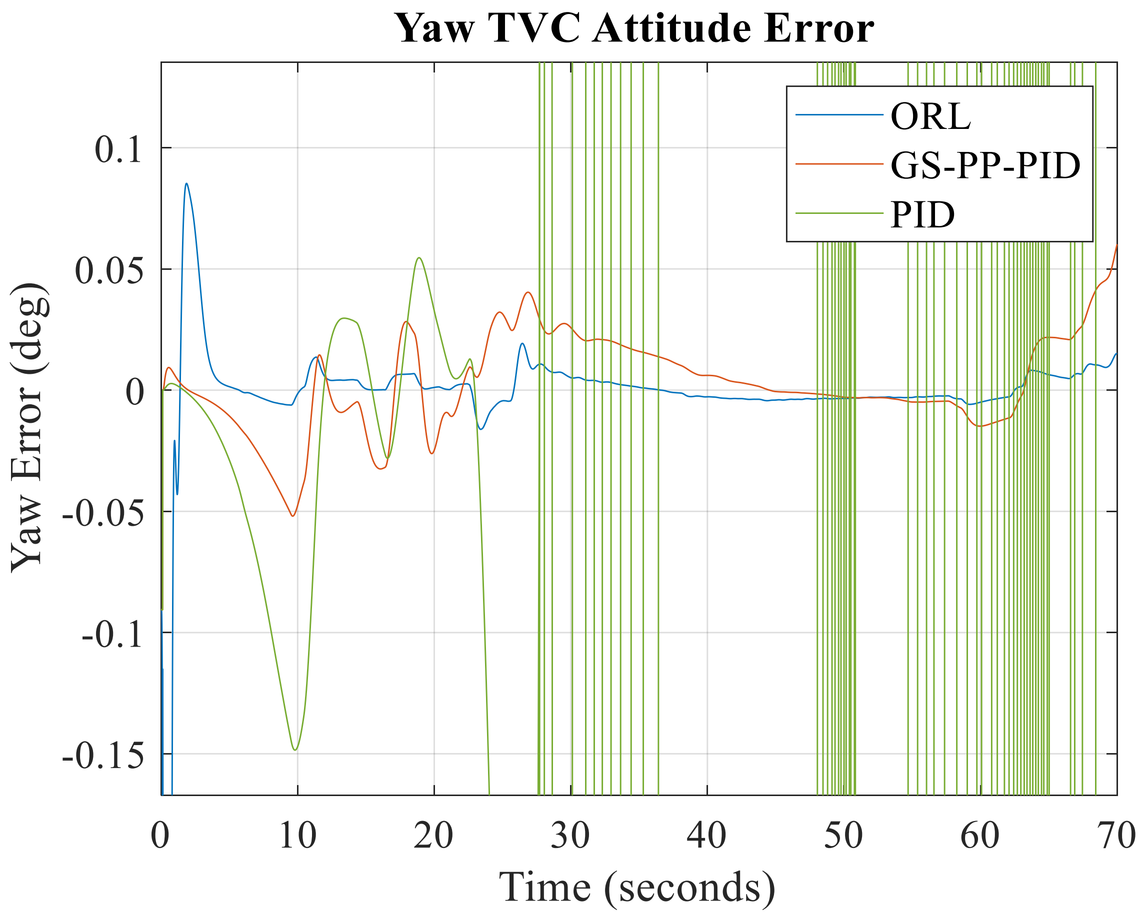
Case 1 Attitude Error



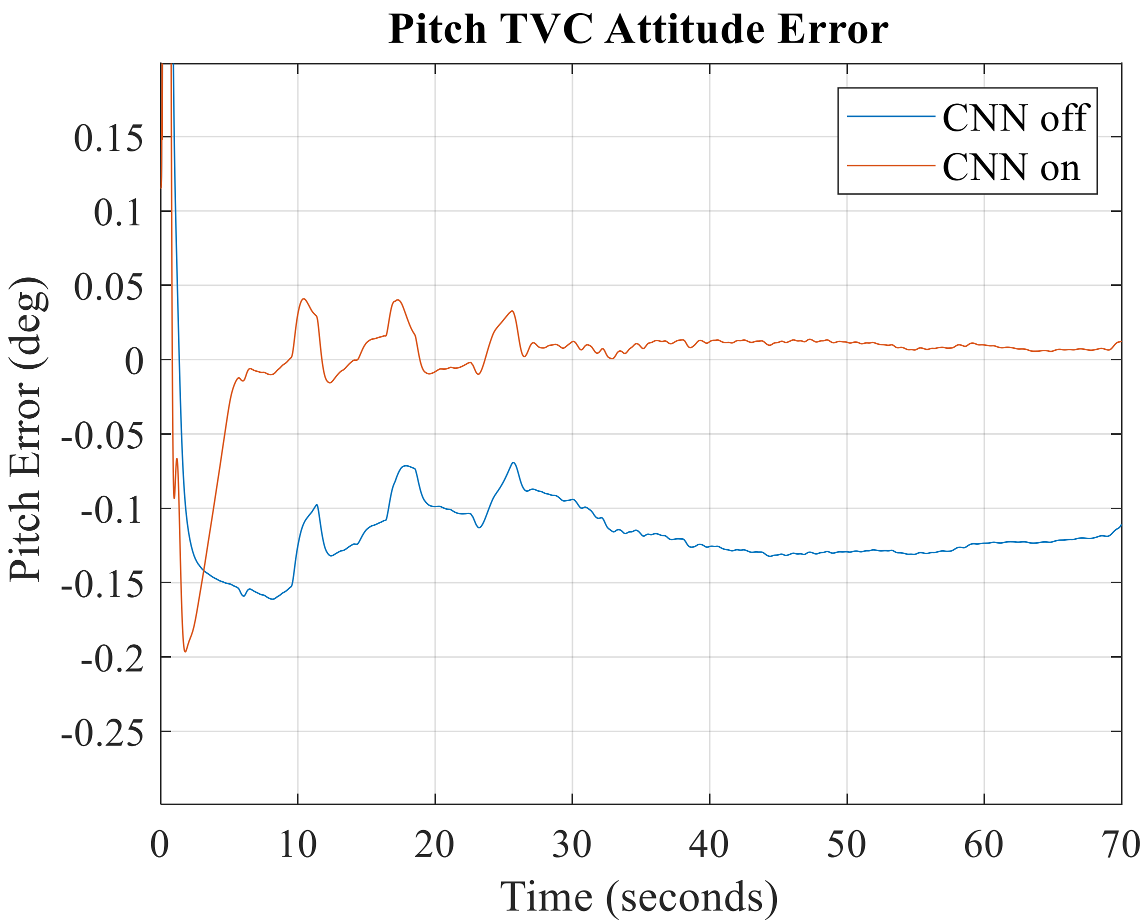


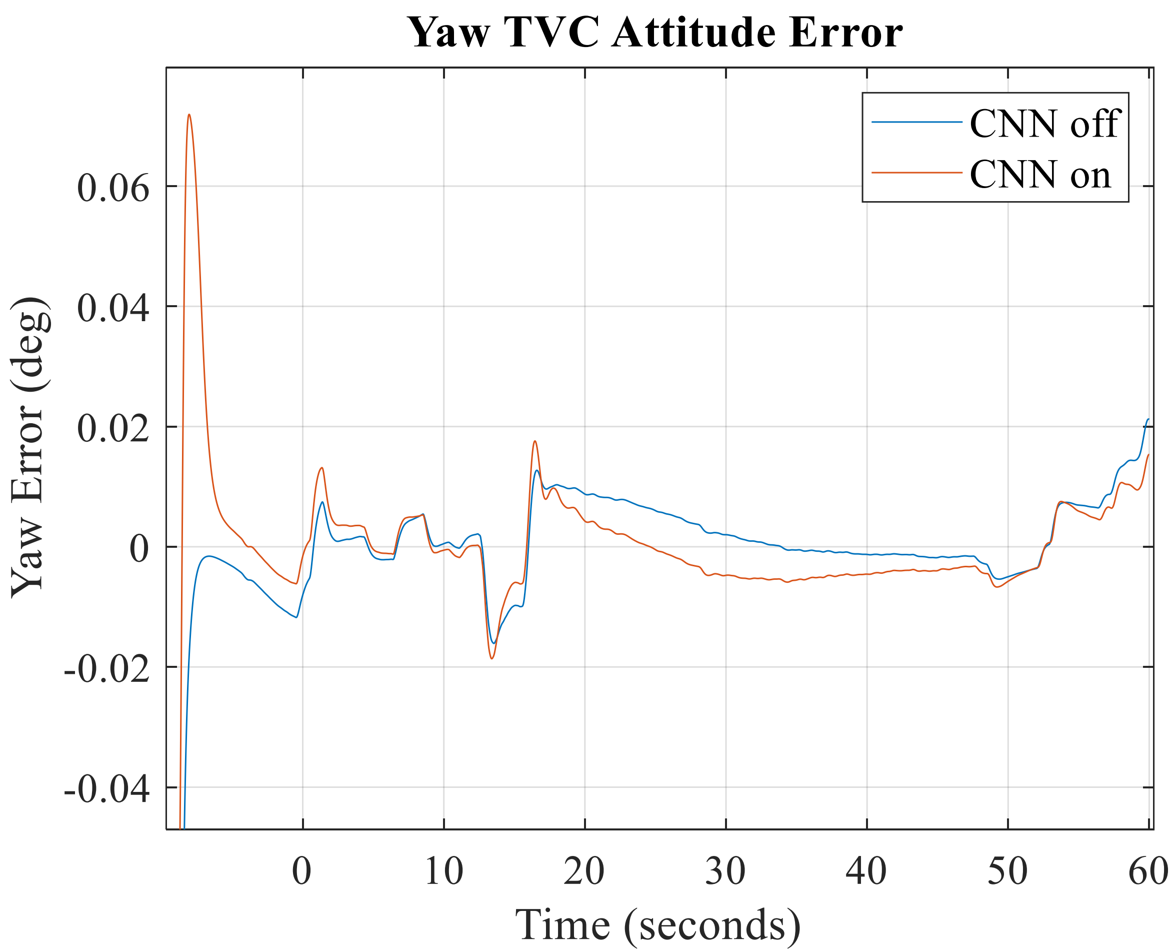
Case 2 Attitude Error



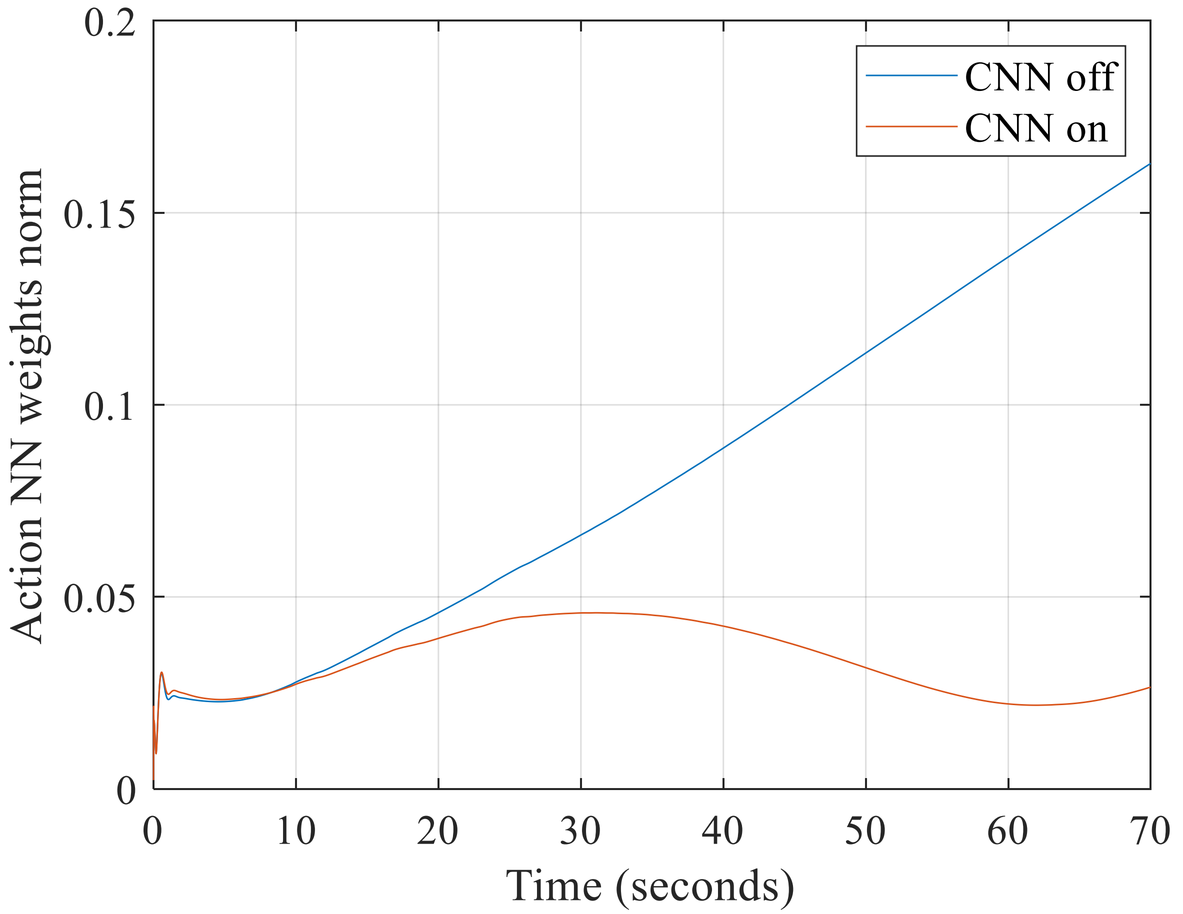


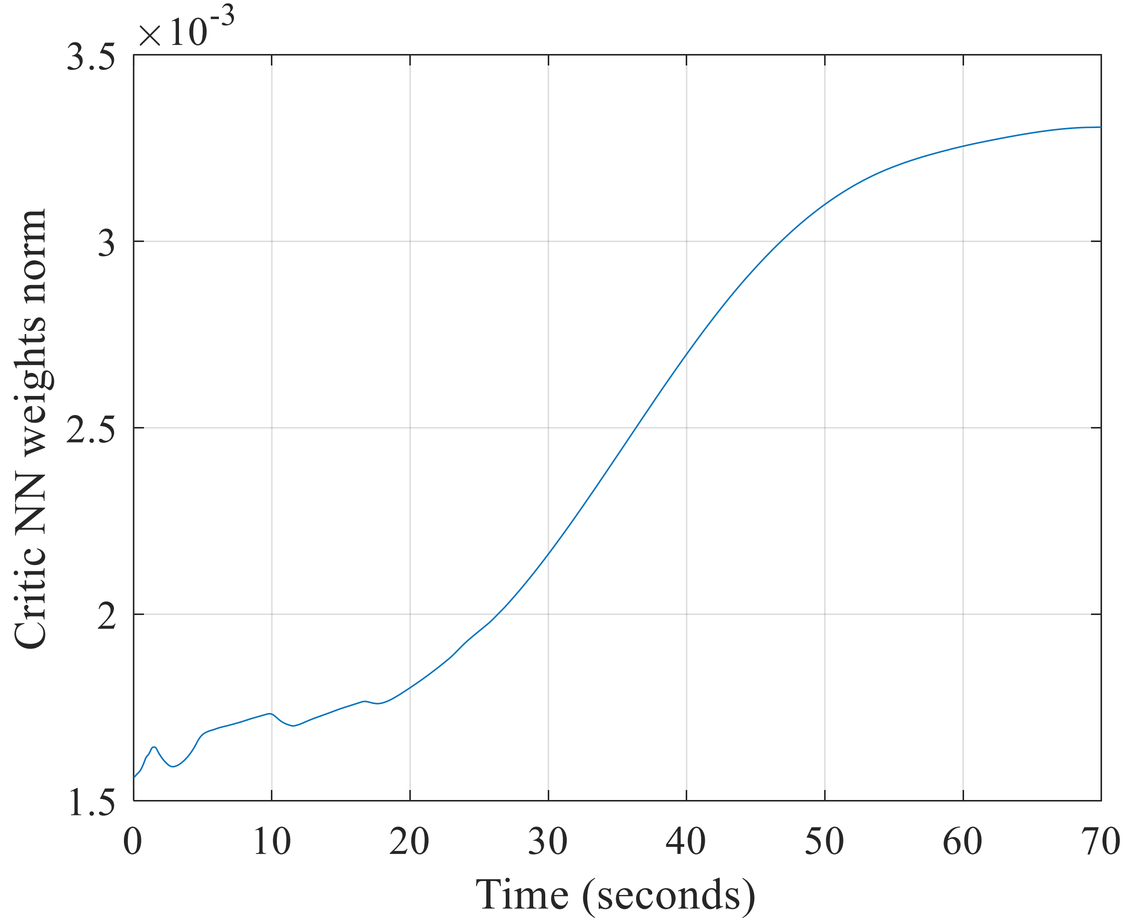
Case 2 Attitude Error with Critic NN switches



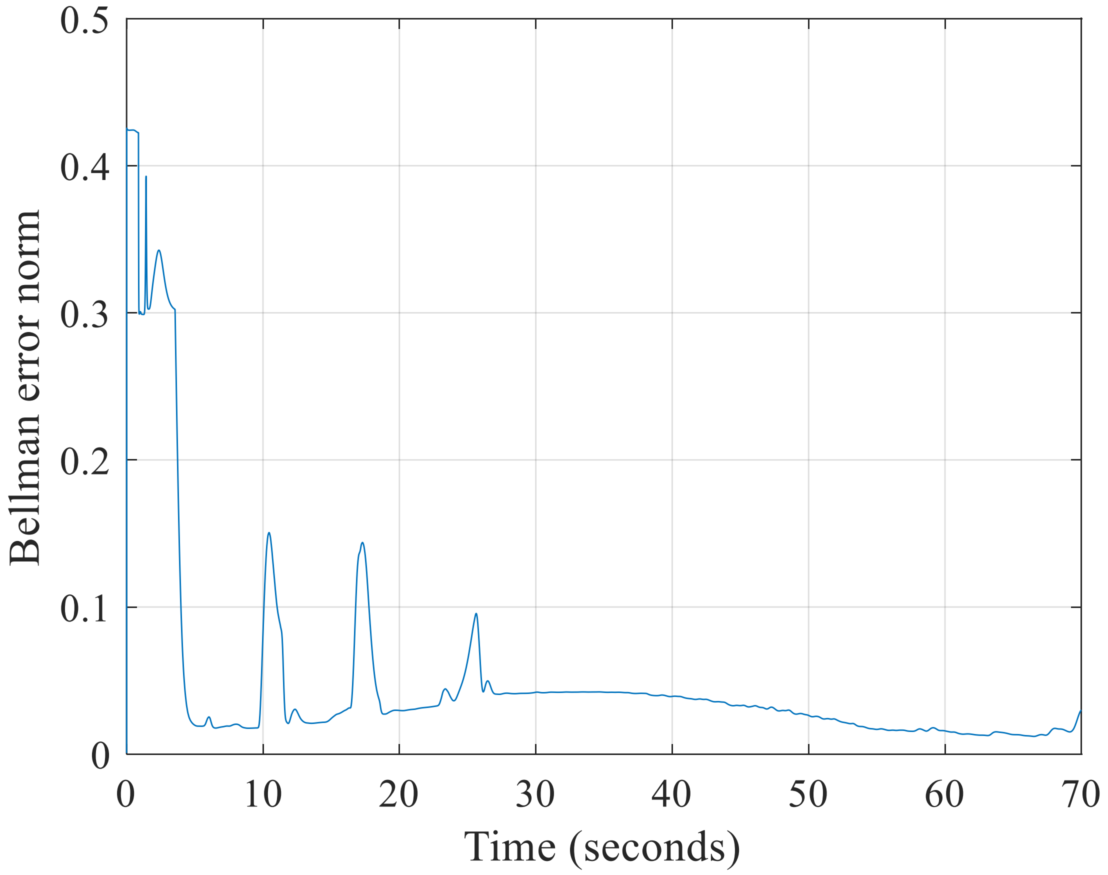


Case 2 Action and Critic Normed NN





Case 2 Normed Bellman Error



Conclusion/Discussion

- The ORL controller has demonstrated superior performance over two other controllers in the following areas:
  - Robustness
  - Adaptability
  - Convergence
- Future works include:
  - Mitigating high transients seen in the attitude errors at the beginning of flight.
  - Assessing the performance under multiple faulty conditions.
  - Extending ORL to inner and outer loops for an integrated optimal solution.

Questions?

Email: hanwoong.bae@nasa.gov

Reference

[1] Yaghoubi, Schnell. “Mars Ascent Vehicle Solid Propulsion Configuration”, 2020 IEEE Aerospace Conference, March 2020

[2] D. Izzo, M. Märtens, and B. Pan, A survey on artificial intelligence trends in spacecraft guidance dynamics and control. Astrodyn. Vol. 3, No. 4, 2019, pp. 287–299.

[3] Q. Yang and S. Jagannathan. Reinforcement learning controller design for affine nonlinear discrete-time systems using online approximators. IEEE Trans Syst Man Cybern B Cybern, 42(2):377–390, April 2012.

[4] Dorf, Richard C., and Robert H. Bishop. *Modern Control Systems*. Upper Saddle River, NJ: Prentice Hall, 2010.